

**ABSTRACT OF THE DISCLOSURE**

A method and apparatus are provided for managing work granules being executed in parallel. A task is evenly divided between a number of work granules. The number of work granules falls between a threshold minimum and a threshold maximum. The threshold minimum and maximum may be configured to balance a variety of efficiency factors affected by the number of work granules, including workload skew and overhead incurred in managing larger number of work granules. Work granules are distributed to processes on nodes according to which of the nodes, if any, may execute the work granule efficiently. A variety of factors may be used to determine where a work granule may be performed efficiently, including whether data accessed during the execution of a work granule may be locally accessed by a node.